

Related URLs - ECSE-88

Author(s): Simon Cantrell

Description of Problem

- The Related URL set of fields occurs multiple places in the UMM-C schema.
- Related URLs are pointers to information to allow the science user to: compute, analyze, visualize information and acquire data.
- Need up-to-date ancillary information, e.g. pointers to documentation on the data set, contact information for the data set engineer, or the responsible science lead. Also need information on how others have used the data set, or if there are known problems.
- UMM, MMT, and CMR changes needed to improve the usage of Related URL information in collection metadata.
- Ideally, once a user has found the collection, we need to minimize the number of steps needed for the user to navigate away from the search results in order to discover the information they require.

JIRA Linkage

~~ECSE-88~~ - Revise Related URL content and usage in UMM, MMT, and CMR **COMPLETED**

Background

Related URLs - this element describes any data/service related URLs that include project home pages, services, related data archives / servers, metadata extensions, direct links to online software packages, web mapping services, links to images, or other data.

Also, a URL associated with the responsible party, e.g. the home page for the DAAC which is responsible for the collection.

Represents Internet sites that contain information related to the data, as well as, related Internet sites such as project home pages, related data archives / servers, metadata extensions, online software packages, web mapping services, and calibration/validation data.

Approach

For UMM and CMR, we propose the following fundamental tenets to guide the resolution of issues related to additional attributes and acquisition information:

- *Focus on the CMR community and their metadata uses.* Metadata standards, recommendations, and guidance only make sense in the context of a community and their use cases for the metadata. When deciding between options for a UMM metadata structure, we choose the one that makes sense for the CMR community. We learn from other communities, such as the ISO community, but our focus is on the metadata needs of the CMR community.
- *Learn from ISO, but don't blindly reuse ISO metadata structures.* Whenever possible, have a single UMM metadata structure that might map to multiple locations in ISO. We don't need the flexibility and complexity of ISO to support the rich metadata needs of the CMR.
- *Whenever possible, specific typed metadata structures should be created from the classification of additional attributes.* For example, in UMM-C-136 our approach would move information from additional attributes to enhance the structure of Platform/Instrument characteristics instead of moving the characteristics into new additional attributes. Over time, this tenet will produce a richer UMM structure that evolves as our understanding of metadata, applications, and missions evolves.

Analysis/Recommendations

Discussion Topics

1. What are the set of Related URLs (i.e. to be found in ECHO, GCMD)?

The most obvious Related URLs are to be found in relation to the following entities:

- a. Citations (e.g. data set reference to show common ways to compute, analyze, visualize or acquire data), Publication References (e.g. location of science papers, or about data set usage), Organizations (e.g. organizations responsible for archive, distribution), Distribution Information (e.g. formats, file sizes, mime types, units), Personnel (e.g. responsible science lead, data set engineer)

2. What are the types of Related URLs?

The best place to start is the Related URL keywords. GCMD keywords are organized by Type and Subtype:

- a. Current GCMD "Type" keywords include: GET DATA, GET RELATED DATA SET METADATA (DIF), GET RELATED DATA SET (SERF), GET RELATED VISUALIZATION, GET SERVICE, VIEW DATASET LANDING PAGE, VIEW EXTENDED METADATA, VIEW PROFESSIONAL HOME PAGE, VIEW RELATED INFORMATION
- b. Current GCMD "Subtype" keywords for "Type" GET DATA include: DATACAST URL, EARTHDATA SEARCH, ECHO, EOSDIS

DATAPOOL, GDS, GIOVANNI, KML, LAADS, LANCE, LAS, MIRADOR, MODAPS, NOAA CLASS, ON-LINE ARCHIVE, OPENDAP DATA, OPENDAP DATA (DODS), OPENDAP DIRECTORY (DODS), REVERB, SSW, SUBSETTER, THREDDS CATALOG, THREDDS DATA, THREDDS DIRECTORY

- c. Current GCMD "Subtype" keywords for "Type" GET RELATED VISUALIZATION include: GIOVANNI
 - d. Current GCMD "Subtype" keywords for "Type" VIEW RELATED INFORMATION include: ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD), GENERAL DOCUMENTATION, HOW-TO, PI DOCUMENTATION, PRODUCT HISTORY, PUBLICATIONS, USER'S GUIDE
 - e. Current GCMD "Subtype" keywords for "Type" GET SERVICE include: ACCESS MAP VIEWER, ACCESS MOBILE APP, GET MAP SERVICE, GET SOFTWARE PACKAGE, GET WEB COVERAGE SERVICE (WCS), GET WEB FEATURE SERVICE (WFS), GET WEB MAP FOR TIME SERVICES, GET WEB MAP SERVICE (WMS), GET WORKFLOW (SERVICE CHAIN), Open Search.
3. We need to more clearly establish relationships which exist between the collection, the URLs and better define their types.
This can be done easily by identifying groups within the GCMD keywords and identify the main types of information to be found at the URLs (Citation, Publication References, Organizations, Distribution Information, Personnel.)
 4. Note that 'Relation' maps to Type in ECHO 10, i.e.,
 - a. <OnlineResource>
<URL> element to supply the base url.
<Type> and,
 - b. 'Relation maps to Content Type, Type and Subtype in DIF, i.e.,
<Related_URL>
<URL> element to supply the base url.
<URL_Content_Type>
<Type>
<Subtype>
 5. Use-case analysis:
If a user has found a collection, he or she will want to either:
 - i) Read about it to find out detailed information, e.g. General Information, Algorithm Theoretical Basis Document, Product History, How-To Guides
 - ii) Get a visual representation of the collection (one or more granules from an orbit, or repeated over a given spatial region, or over intervals of time), e.g. Browse Image (JPEG, PNG, TIFF, GeoTIFF), GIS Mapping Representation (WCS, WFS, WMS)
 - iii) Get the data from the collection (one or more granules from an orbit, or repeated over a given spatial region, or over intervals of time), e.g. Native, ASCII, Binary, HDF4, HDF5, HDF-EOS, KML, etc. or consume the data via a service, e.g. OPENDAP, DODS, THREDDS, or via one of the GIS Mapping services (need auto-discovery of OPENDAP, etc. endpoints)
 - iv) Perform scientific analysis on the data, e.g using one of a number of standard analysis tools, or custom tools (Panoply, GRADs, IDL, Matlab, Ferret), or HDFView, IDV or Cal/Val tools (i.e. JPSS GRAVITE, Apache OODT)
 - v) Contact people or organizations which can provide further information about the data, e.g. Principal Investigator, Data Set Engineer.

Note: We need to distinguish between the different types of user, i.e. Data Provider, Science User, Cal/Val Data Quality Engineer, Emergency Responder. This will be done in the Use Cases, which will be the topic of separate ticket.

The sections below detail the proposed changes based on the analysis done by the ECSE team:

Changes to UMM-C fields

PublicationReference

Currently:

Change to:

PublicationReference	
/PublicationReference/RelatedUrl	/PublicationReference/RelatedUrl
/PublicationReference/RelatedUrl/ContentType	/PublicationReference/RelatedUrl/UrlContentType
/PublicationReference/RelatedUrl/Relation	
/PublicationReference/RelatedUrl/ContentType/Type	/PublicationReference/RelatedUrl/Type

/PublicationReference/RelatedUrl/ContentType/Subtype	/PublicationReference/RelatedUrl/Subtype
/PublicationReference/RelatedUrl/Description	/PublicationReference/RelatedUrl/Description
/PublicationReference/RelatedUrl/URL	/PublicationReference/RelatedUrl/URL

Recommend using UrlContentType as a field in its own right to describe the relationship between the entity (in this case: PublicationReference) and the URL. Use a combination of Type and Subtype to group and organize the URLs. Repeat this method to describe the relationships for all other entities in the collection, which have Related URLs (i.e. Distribution, Organization (now DataCenter, Personnel (now DataContact)).

Distribution

Currently:

Change to:

<u>Distribution</u>	
/Distribution_Information/Related_URL	/Distribution/RelatedUrl
/Distribution_Information/Related_URL/ContentType	/Distribution/RelatedUrl/UrlContentType
/Distribution_Information/Related_URL/ContentType/Type	/Distribution/RelatedUrl/Type
/Distribution_Information/Related_URL/ContentType/Subtype	/Distribution/RelatedUrl/Subtype
/Distribution_Information/Related_URL/Description	/Distribution/RelatedUrl/Description
/Distribution_Information/Related_URL/FileSize	
/Distribution_Information/Related_URL/MimeType	/Distribution/RelatedUrl/GetService/MimeType
/Distribution_Information/Related_URL/Protocol	/Distribution/RelatedUrl/GetService/Protocol
/Distribution_Information/Related_URL/Title	
/Distribution_Information/Related_URL/URL	/Distribution/RelatedUrl/URL
/Distribution/Media	/Distribution/RelatedUrl/GetData
/Distribution/FileSize	
/Distribution/Size	/Distribution/RelatedUrl/GetData/Size
/Distribution/Unit	/Distribution/RelatedUrl/GetData/Unit
/Distribution/Format	/Distribution/RelatedUrl/GetData/Format
/Distribution/Fees	/Distribution/RelatedUrl/GetData/Fees
	/Distribution/RelatedUrl/GetService
	/Distribution/RelatedUrl/GetService/FullName
	/Distribution/RelatedUrl/GetService/DataID
	/Distribution/RelatedUrl/GetService/URI

Recommend sifting out the metadata fields which support distribution and sort them into those that support getting the data via a simple download, and those which support getting the data via a service, e.g. via OPeNDAP, etc. Create two new classes to group these attributes: GetData and GetService. Model the RelatedUrl class in a common way across the UMM, and remove any unnecessary attributes.

Organization

Currently:

Change to:

<u>Organization</u>	
/Organization/Party/RelatedUrl	/DataCenter/ContactInformation/RelatedUrl
/Organization/Party/RelatedUrl/ContentType	/DataCenter/ContactInformation/RelatedUrl/UrlContentType
/Organization/Party/RelatedUrl/Relation	
/Organization/Party/RelatedUrl/ContentType/Type	/DataCenter/ContactInformation/RelatedUrl/Type
/Organization/Party/RelatedUrl/ContentType/Subtype	/DataCenter/ContactInformation/RelatedUrl/Subtype
/Organization/Party/RelatedUrl/Protocol	
/Organization/Party/RelatedUrl/Title	
/Organization/Party/RelatedUrl/Description	/DataCenter/ContactInformation/RelatedUrl/Description
/Organization/Party/RelatedUrl/Caption	
/Organization/Party/RelatedUrl/URL	/DataCenter/ContactInformation/RelatedUrl/URL
/Organization/Party/RelatedUrl/MimeType	
/Organization/Party/RelatedUrl/FileSize	
/Organization/Party/RelatedUrl/FileSize/Size	
/Organization/Party/RelatedUrl/FileSize/Unit	

Updated to reflect most recent changes in Organization and Personnel. DataCenter has a ContactInformation which has its RelatedURL. Similarly, DataContact has a ContactInformation which has its RelatedURL, as shown below.

Personnel

Currently:

Change to:

<u>Personnel</u>	
/Personnel/Party/Person /RelatedURL	/DataContact/ContactInformation/RelatedUrl
/Personnel/Party/Person /RelatedURL/ContentType	/DataContact/ContactInformation/RelatedUrl/UrlContentType
/Personnel/Party/Person /RelatedURL/Relationship	
/Personnel/Party/Person /RelatedURL/ContentType/Type	/DataContact/ContactInformation/RelatedUrl/Type

/Personnel/Party/Person RelatedURL/ContentType/Subtype	/DataContact/ContactInformation/RelatedUrl/Subtype
/Personnel/Party/RelatedUrl/Protocol	
Personnel/Party/RelatedUrl/Title	
Personnel/Party/RelatedUrl/Description	/DataContact/ContactInformation/RelatedUrl/Description
Personnel/Party/RelatedUrl/Caption	
Personnel/Party/RelatedUrl/URL	/DataContact/ContactInformation/RelatedUrl/URL
Personnel/Party/RelatedUrl/MimeType	
Personnel/Party/RelatedUrl/FileSize	
Personnel/Party/RelatedUrl/FileSize/Size	
Personnel/Party/RelatedUrl/FileSize/Unit	

Changes to UMM-Common fields

Publication

Currently:

Change To:

<u>Publication</u>	
/ResourceCitationType	/ResourceCitationType
/ResourceCitationType/Version	/ResourceCitationType/Version
/ResourceCitationType/RelatedUrl	/ResourceCitationType/OnlineResources
/ResourceCitationType/Title	/ResourceCitationType/Title
/ResourceCitationType/Creator	/ResourceCitationType/Creator
/ResourceCitationType/Editor	/ResourceCitationType/Editor
/ResourceCitationType/SeriesName	/ResourceCitationType/SeriesName
/ResourceCitationType/ReleaseDate	/ResourceCitationType/ReleaseDate
/ResourceCitationType/ReleasePlace	/ResourceCitationType/ReleasePlace
/ResourceCitationType/Publisher	/ResourceCitationType/Publisher
/ResourceCitationType/IssueIdentification	/ResourceCitationType/IssueIdentification
/ResourceCitationType/DataPresentationForm	/ResourceCitationType/DataPresentationForm
/ResourceCitationType/OtherCitationDetails	/ResourceCitationType/OtherCitationDetails
/ResourceCitationType/DOI	/ResourceCitationType/DOI

Distribution

Currently:

Change to:

Distribution	
/RelatedUrlType	/RelatedUrl/UrlContentType
	/RelatedUrl/Type
	/RelatedUrl/Subtype
/RelatedUrlType/URLs	/RelatedUrl/URL
/RelatedUrlType/Description	/RelatedUrl/Description
/RelatedUrlType/Relation	
/RelatedUrlType/MimeType	/Distribution/RelatedUrl/GetService/MimeType
/RelatedUrlType/FileSize	/Distribution/RelatedUrl/GetData/Size

Recommend removing the Relation field entirely. URL relationships can be described using UrlContentType attribute. URLs can be grouped by type and subtype, for each UrlContentType.

Currently:

Change to:

FileSizeType	
/FileSizeType	
/FileSizeType/FileSize	/RelatedUrl/GetData
/FileSizeType/Size	/RelatedUrl/GetData/Size
/FileSizeType/Unit	/RelatedUrl/GetData/Unit

and (consider both of these sets of attributes together) - now members of the new GetData class.

Currently:

Change to:

DistributionType	
/DistributionType/	
/DistributionType/DistributionMedia	
/DistributionType/Sizes	
/DistributionType/DistributionFormat	/RelatedUrl/GetData/Format
/DistributionType/Fees	/RelatedUrl/GetData/Fees

Need to reconcile any URLs in the UMM-C and UMM-Common data models and determine which RelatedUrl class carries the most weight and remove the other class.

Mappings to DIF, ECHO, ISO

Publication

DIF 9	DIF 10	ECHO 10	ISO	Update UMM-C to:
/DIF/Reference/Online_Resource	/DIF/Reference/Online_Resource		CI_OnlineResource	OnlineResource
	/DIF/Reference/Citation			

Recommend incorporating the OnlineResource class to the UMM as an alternate to RelatedURL. Useful for Web pages which contain resources, e.g. downloadable files (e.g. CSV, PDF, DOC files).

Use ISO CI_OnlineResource attributes: Linkage, Name, Description, Function at a minimum. See UMM-Common model diagram for class specification.

Organization

DIF 9	DIF 10	ECHO 10	ISO	Update UMM-C to:
Data_Center/Data_Center_URL	Organization/Organization_URL			Organization/RelatedUrl

Recommend using the RelatedUrl class for Organization URLs.

Distribution (Browse)

DIF 9	DIF 10	ECHO 10	ISO	Update UMM-C to:
/DIF/Multimedia_Sample	/DIF/Multimedia_Sample	/Collection/AssociatedBrowsingImageURLs		RelatedUrl

Recommend using the RelatedUrl class for browse image, or multimedia sample distribution. Differentiate between browse and native media distribution with UrlContentType, Type and Subtype attributes. Example" UrlContentType = <DistributionURL>, Type = <GetData>, Subtype = <Browse> or <Media> depending on whether the file for distribution is a browse file or a native media file.

Distribution (Media)

DIF 9	DIF 10	ECHO 10	ISO	Update to UMM-C:
/DIF/Related_URL/URL_Content_Type	/DIF/Related_URL/URL_Content_Type			RelatedUrl/UrlConte
/DIF/Related_URL/URL_Content_Type/Subtype	/DIF/Related_URL/URL_Content_Type/Subtype			RelatedUrl/Subtype
/DIF/Related_URL/URL_Content_Type/Type	/DIF/Related_URL/URL_Content_Type/Type			RelatedUrl/Type
		/Collection/OnlineAccessURLs		RelatedUrl
		/Collection/OnlineAccessURLs /OnlineAccessURL		
		/Collection/OnlineResources		OnlineResource
		/Collection/OnlineResources/ OnlineResourceURL		

Recommend using either RelatedUrl class or OnlineResource class as appropriate. Depending on whether an Related URL or Online Resource URL.

A Complete Mapping will be added via a separate spreadsheet (a marked up version of Erich's excel sheet).

Suggested UMM-C class revisions:

Bring together all information which describes a collection that has some type of on-line resource. For example:

WebResource/OnLineResource (1..*)

WebResource/RelatedUrl (1..*)

Add two new objects to the RelatedUrl Object: GetData and GetService with Relationships: IsDistributedVia

RelatedURL/DistributedVia/GetData (0..*) { Reference to a GetData Object which can distribute the actual data, in a given data format, size and units }

RelatedURL/DistributedVia/GetService (0..*) { Reference to a GetService Object which can visualize, subset, or distribute the actual data, in a given mime type, protocol, DataType, DataID, service, URI }

Suggested DIF/UMM/ECHO mappings are captured ([see Notes column for changes in Green](#)) in the attached spreadsheet (DIF-UMM-ECHO_Mapping_06102016.xlsx). UMM-C (Modified) column shows added/removed class/attributes.

Suggested CMR-Common UML Model

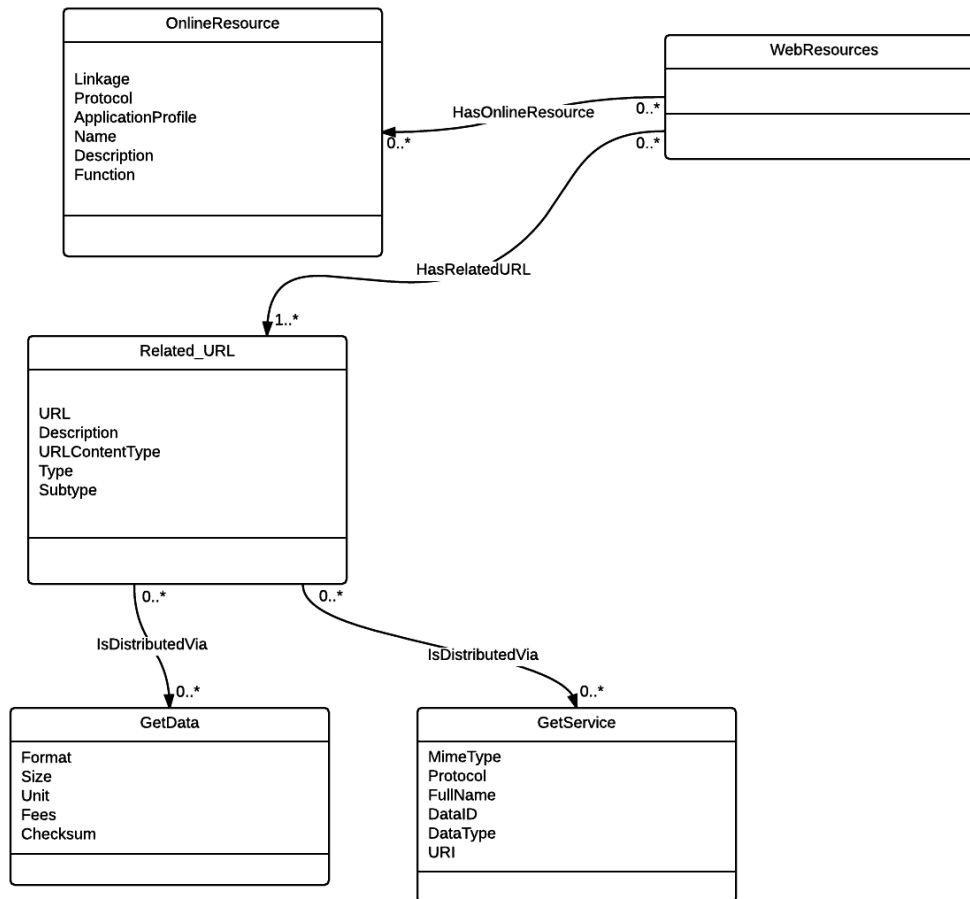


Figure 1. RelatedURL (Modified) UML Model

For RelatedURL the attributes are:

/RelatedURL/URL <The URL, e.g. "<https://daac.ornl.gov/MODIS/>">

/RelatedURL/Description <This is the description of the URL, e.g. "The FTP base directory for the collection">

/RelatedURL/URLContentType <valid keywords are: "CollectionURL", "PublicationURL", "OrganizationURL", "DistributionURL", "PersonnelURL", "VisualizationURL">

/Related/URL/Type <valid keywords are: "GetData", "GetService", or "GetVisualization">

/RelatedURL/SubTypes <valid keywords for GetData are: "Browse", "Media"> also for GetService: <"Web Coverage Service", "Web Feature Service", "Web Mapping Service", "OPENDAP", "THREDDS", "DODS", also for GetVisualization: "MIRADOR">.

Recommend reviewing the use of Type and Subtype in the relevant GCMD keyword list for consistency of usage, and adding the URLContentType keywords to enable the definition of a more comprehensive, unambiguous RelatedURL set of metadata.

For GetData object class

Add the following attributes:

/GetData/Format

/GetData/Size

/GetData/Description

/GetData/Unit

/GetData/Fees

/GetData/Checksum

(This is the minimum set of attributes to enable the client to support the download of a data set).

For GetService object class

Add the following attributes:

/GetService/Protocol

/GetService/MimeType

/GetService/FullName

/GetService/DataID

/GetService/DataType

/GetService/URI (0..*) URI(s) of actual data - as distinct from Catalog URL - use /RelatedUrl/URL for root catalog entry, e.g, for a THREDDS service.

(This is the minimum set of attributes to enable the client to support visualization, subsetting or download of a data set via a service).

Depending on the RelatedURL "Type", either <"GetData", GetService", the Earthdata Search client can now distinguish between distribution of data via direct download (media) and/or via services.

With suitable additional software on the client-side, the data can either be acquired directly in the appropriate format, or displayed in the client via a suitable service. The services typically provide metadata to aid auto-discovery. This can be the subject of a separate ticket.

Interoperability Considerations

We are not adopting any ISO classes or attributes directly into the RelatedURL class. However, we are proposing to include the OnlineResource class, which maps directly to the CI_OnlineResource class.

RelatedURL class will be used to expand and clarify URL content types, types, and subtypes. It will ease mapping issues to the current DIF9, DIF10 and ECHO10 fields for most URLs. For some URLs, specifically those associated with Citations, OnlineResource class provides an alternate to RelatedURL and helps to bring the CMR towards interoperability with other systems.